

A supervised method for unbiased peer-to-peer evaluation. An experience with engineering students

Simois, Francisco J.*

*Department of Signal Theory and Communications, University of Sevilla
Engineering School, Camino de los Descubrimientos, s/n, 41092, Sevilla (Spain)*

* Corresponding author: Email: fjsimois@us.es; Phone: +34 954 48 81 33

Received: 2015-02-14; Accepted: 2015-06-25

Abstract

Continuous evaluation is an assessment method which has some appealing advantages but also implies an increase of the teacher's efforts and it may be unfeasible if the class is large.

Of course, new technologies may be used to implement automatized evaluations, but it is usually quite difficult to carry them out when a complex task like an engineering problem is to be judged.

An interesting alternative is a peer-to-peer evaluation, that is, the students themselves review their works. Nevertheless, one drawback is that it is likely that the grades are overrated. Although this is a well-known problem, not much effort is usually put into solving it. In this work we propose a novel method to limit this inconvenience, which is that the teacher randomly supervises a fraction of the students tasks.

In this paper we present the results of such an experience carried out in a Signal Processing course within a Robotics Engineering degree. More precisely, four different sets of problems were solved by the teacher in class. At the same time, they were peer-to-peer reviewed by the students, following the indications given by the professor. Later, when the random supervision is performed, a penalty is applied if a major flaw in a student's evaluation is detected. Thanks to this strategy, the scores tended to be more and more accurate according to the teacher's criteria.

Finally, the results of a survey anonymously fulfilled by the students to assess this experience are also presented.

Keywords

Peer-to-peer evaluation, supervised assessment, engineering course, higher education.

1. Introduction

The Bologna Declaration signed on June 19, 1999, proposes, among other aspects, the convergence of the various European higher education systems. This will help to yield a knowledge-based economy capable of growing sustainably and providing more and better jobs and a higher social cohesion (Bologna, 1999).

The creation of the European Higher Education Area entails a new educational paradigm that enables the development of new educational models and promotes a new skills-based learning to train graduates in solving the problems they will have to address in their jobs (EHEA, 2010). These skills are called competencies. It is considered necessary for students to have competencies for regulating individual and group work, for establishing learning goals, planning courses of action, selecting suitable strategies and resources, and reviewing and reorienting tasks in order to meet predetermined objectives (Torrano, 2004; Pintrich, 2000).

Moreover, the need to use evaluation for pedagogical ends has been highlighted by numerous authors (see, for example, Schunk, 1998; Coll, 1999; William, 2000; Broadfoot, 2004; McDonald, 2006). Continuous assessment is used for purposes of feedback that can serve to improve the learning of the students and to enhance the teaching. This kind of evaluation allows to identify errors in the process, adjusting and orienting it (Delgado, 2005). It has to be performed during the learning process in order to detect the learning gaps.

Continuous assessment promotes the pace of study established by the professor, which involves the study of learning small units distributed over extended periods of time. This perspective emphasizes “assessment for learning” (Birembaum, 2006; Nunziati, 1990; Allal, 1991), that is, the importance of providing students with information about their own learning process, as well as possible ways for improving it.

The process of European convergence has also prompted implementation of teaching methodologies centered on students’ autonomous work (Coll, 2007). That is, the

contemporary context wherein higher education operates has determined that students should play a more active role in the teaching-learning process and, likewise, that teachers should play a facilitator role in that process (Arias, 2014).

In addition, in recent times has emerged a movement called “authentic assessment”, “performance assessment” or “alternative assessment” (Schlichting, 2015; Biggs, 2006; Birembaum, 2006; Díaz Barriga, 2006), which emphasizes methods that facilitate direct observation of student work. In this line, control of evaluation has been transferred from teacher to student. The agent evaluation are not only the teachers, but students play an important role in their own assessment or that of their peers (Mateo, 2005).

This requires that the student assumes the learning objectives and criteria to be used for his evaluation. This formative evaluation is compatible with the continuous one since it considers that the learning is better when targets, criteria and standards are known. Therefore, both autoevaluation and peer-to-peer (between students) reviews are promoted and encouraged.

One well-known issue of peer assessment is its validity (that is, the bias with respect to a professor evaluation) and its reliability (that is, the dispersion of grades with respect to their average). Some authors have carried out an extensive review of this matter (Falchikov 2000). Actually, although some previous works have shown no significant bias in peer assessment (Xiao, 2008; Marks, 2013), many others have found certain problems in its validity (Kommalage, 2011; Harris, 2011) or even very important flaws (De Grez, 2012). Despite of that, previous works seldom propose some ways to reduce the bias and only recommendations about how to manage students’ attitude and responsiveness are provided (Lansiquot, 2015; Koç, 2011).

Starting from this perspective, the objective of the present study is to introduce a new peer-to-peer evaluation methodology performed with engineering students. Due to the intrinsic difficulty of the proposed tasks, overrated grades are expected. The novelty of the proposed method is that, in order to reduce this bias, the professor supervises a

fraction of the marks and applies penalties to the students that try to cheat. Using this approach it is intended that the peer assessment adjusts itself properly during the course, without increasing too much the professor's effort.

2. Methods

In the next paragraphs, the proposed methodology will be presented. First, we will describe the current situation of the subject under study. Second, we will show the details of the learning experience.

2.1 Current situation and problem statement

The subject in our study is “Procesamiento Digital de la Señal” (Digital Signal Processing). It is taught in the third course of the Electronics, Robotics and Mechatronics Engineering degree at the University of Sevilla. The course has an applied nature, having several laboratory sessions. In addition, it is intended that the students try to solve a considerable amount of applied problems to reinforce the theoretical ideas that are taught. These problems are fully graded because they are one important item in the continuous evaluation.

There was 49 students participating actively in the course and only one professor. Taking into account that the total number of problems to be reviewed is 58, this implies a huge amount of effort and time for the teacher. Moreover, the intrinsic nature of the problems is heterogeneous and complex; in addition, the grading criteria is not merely correct/incorrect but they are evaluated as a whole. Therefore, an automatic evaluation tool is almost impossible to be performed.

For all those reasons, a peer-to-peer assessment was carried out. That is, the students themselves had to grade their classmates. The issue is that it is very likely that the scores

are overrated in such a situation. To overcome this problem, a supervised evaluation is proposed. The details are as follows.

2.1.1 The supervised peer-to-peer evaluation

There are four different sets of problems, each of them intended to be solved by the professor in a different session. At the beginning of the session, each student freely chooses a classmate and they swap their problems in order to be graded. The only restriction is that nobody can review the same classmate more than once.

When everybody has exchanged his problems, the professor solves them and, at the same time, provides the criteria for grading. These criteria are as detailed as possible in order to the students to award clear and impartial marks. In addition, the teacher answers the questions about grading that the students may pose on the fly.

At the end of the class, the students fill a form with three fields (their name, the name of whom they have graded and the marks awarded) and give it to the professor. Finally, the teacher chooses randomly 12 students and collects their sets of problems. To be precise, there is one restriction in the randomness: every student will have to be selected at least once and at most three times during the course.

Later, the professor reviews the 12 students he have chosen and compares his grades to the peer-to-peer ones. If a discrepancy of a 10% or more is found, a penalty of twice the difference is applied to the evaluator. For instance, let us assume that student X grades student Y with a 6 and the teacher with a 5 (that is, a difference of 1 point or a $1/6 \cdot 100 = 16.7\%$). Then, the mark of student X (not student Y) is diminished in $2 \cdot 1 = 2$ points. Of course, the students were previously warned about this issue.

Finally, the professor publishes the final grades with all the associated details. The final mark is the teacher's one, if it exists (since it prevails over the peer-to-peer evaluation),

minus the penalty; or the classmate's grade, if the professor's one does not exist, minus the penalty.

On the other hand, an anonymous survey was conducted at the end of the course. The purposes were mainly two: assess whether the methodology had helped the students with the subject; and whether the action of grading itself had been puzzling to them.

3. Results and discussion

On the one hand, we will present the evolution of the peer evaluation compared to the professor's one along the course. On the second hand, the results of the satisfaction survey will be shown.

3.1. Comparison between peer-to-peer and professor's evaluations

As it has been previously explained, two different evaluations were carried out for each set of problems. First of all, all students were graded by a classmate. Secondly, some randomly selected students were also reviewed by the professor. Both marks would be the same in the ideal case but it is unavoidable that some differences appear.

These differences are clearly shown in Figure 1. In (a), it is seen that, for the first set of problems, the grades awarded by the classmates were as much as 0.5 points higher than the grades given by the professor. This seems to indicate that, despite the fact that the teacher cautioned the students about the possibility that a sanction was applied, they did not consider it a real warning. Probably, many of them thought that the professor was only threatening but he would not apply the penalty; or perhaps they were overconfident that they would not be chosen to be revised.

As a consequence, some penalties were imposed. Actually, 4 out of 12 students graded by the professor were penalized. To be precise, if a positive difference of a 10% of more between the two evaluations was found, a penalty of twice the difference was applied to the involved student. The details are shown in Table 1.

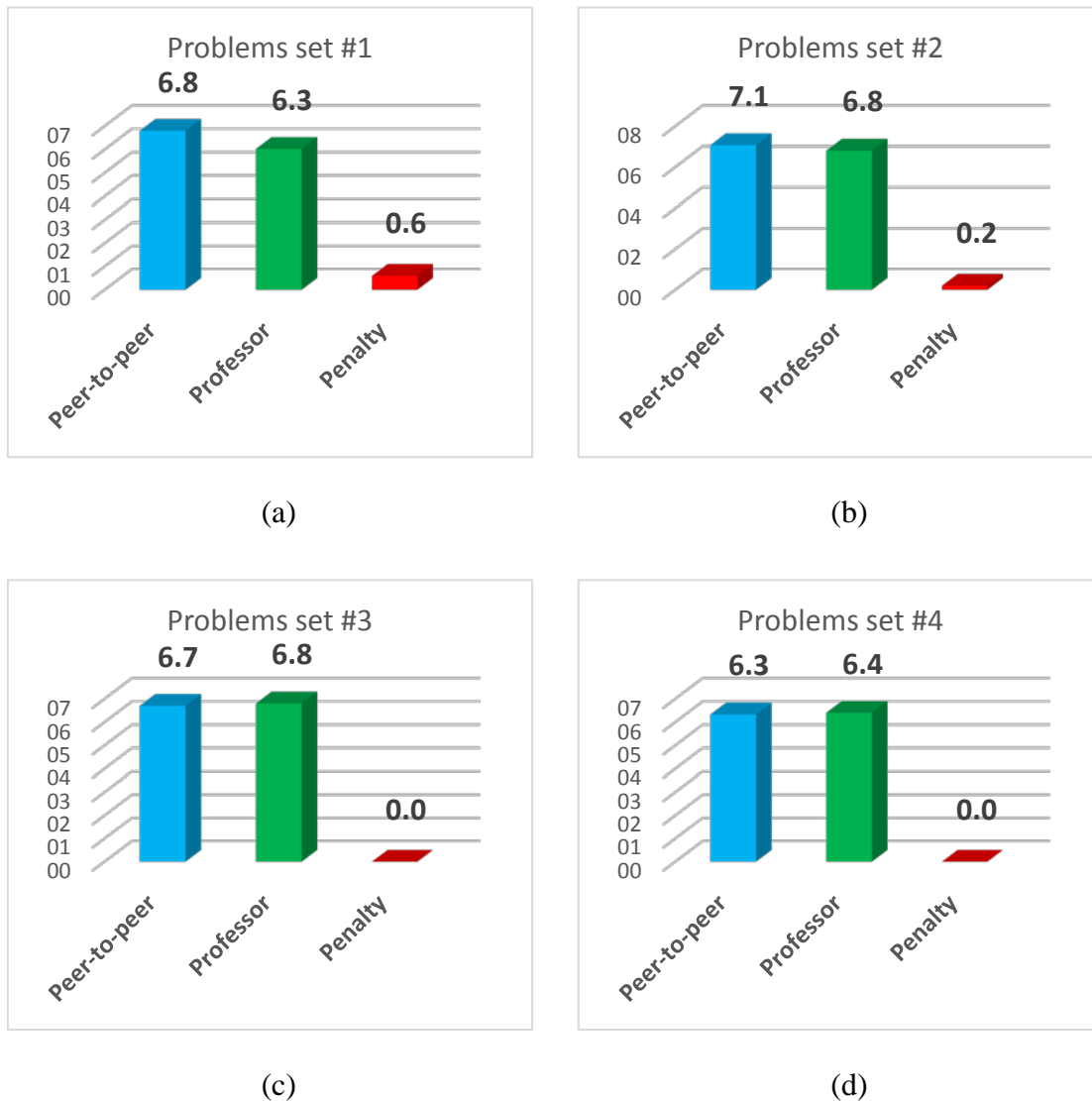


Figure 1. Grades awarded by both the peer-to-peer and the professor's evaluation, as well as the penalty applied for (a) first set of problems, (b) second set of problems, (c) third set of problems and (d) fourth set of problems.

Table 1. Details of the grades for the first set of problems.

	Peer-to-peer	Professor	Difference	Difference (%)	Penalty
Student #1	7.4	6.8	+0.6	+8.1%	0.0
Student #2	7.0	5.9	+1.1	+15.7%	2.2
Student #3	9.1	9.3	-0.2	-2.2%	0.0

Student #4	7.3	6.8	+0.5	+6.8%	0.0
Student #5	7.2	6.2	+1.0	+13.9%	2.0
Student #6	5.3	4.4	+0.9	+17.0%	1.8
Student #7	6.1	5.5	+0.6	+9.8%	0.0
Student #8	6.8	7.0	-0.2	-2.9%	0.0
Student #9	7.2	6.7	+0.5	+6.9%	0.0
Student #10	3.8	3.0	+0.8	+21.1%	1.6
Student #11	8.8	8.8	0.0	0.0%	0.0
Student #12	5.9	5.6	+0.3	+4.3%	0.0
Average	6.8	6.3	+0.5	+8.2%	0.6

Of course, it is apparent that, in general, the peer-to-peer grades tends to be overrated, which is not surprising. Specifically, there is an average of an extra 8.2% in the marks awarded by the classmates.

In Figure 1 (b) the situation is quite different. Although there is a still a slight tendency to overrating (details are omitted), it is much lower than in (a). Indeed, only 1 out of 12 students was penalized. Obviously, they realized that the warning was for real and the majority of them did not risked again.

Finally, in Figure 1 (c) and (d) the situation is completely stabilized. All students followed precisely the criteria given by the professor when the problems were solved in class. Actually, although some subtle differences between the peer-to-peer and the teacher's evaluation were present, no penalty was applied and indeed the average grade awarded by the classmates was slightly lower than the professor's one.

In conclusion, it can be stated that the main goal of this experience was fully achieved. That is, the peer evaluation was finally free of bias and this was done even taking into account that the professor did actually grade just a fourth of the students.

3.2. Results of the satisfaction survey

At the end of the course, an anonymous satisfaction survey was conducted in order to get the students' opinion on this evaluation methodology. The survey consisted of five questions, which are shown in Table 2. Each question had to be answered with a number from 1 to 5 (see Table 3). The results of the survey are presented in Figure 2.

It can be seen that the great majority of students believe that the peer-to-peer evaluation has helped them to keep the subject up to date. This is a very important point, since the students usually tend to prepare the subject just at the end, at least in courses where the evaluation consists in just a final exam.

Table 2. Questions in the satisfaction survey.

Question #1	Peer evaluation has helped you to keep the subject up to date?
Question #2	Peer evaluation has helped you to better understand the subject?
Question #3	Do you agree with the grades you received from your classmates?
Question #4	It has been difficult to you to grade your classmates?
Question #5	Overall, the peer evaluation has been positive to you?

Table 3. Possible answers to the survey.

1	Totally disagree
2	Disagree
3	Partially agree
4	Agree
5	Totally agree

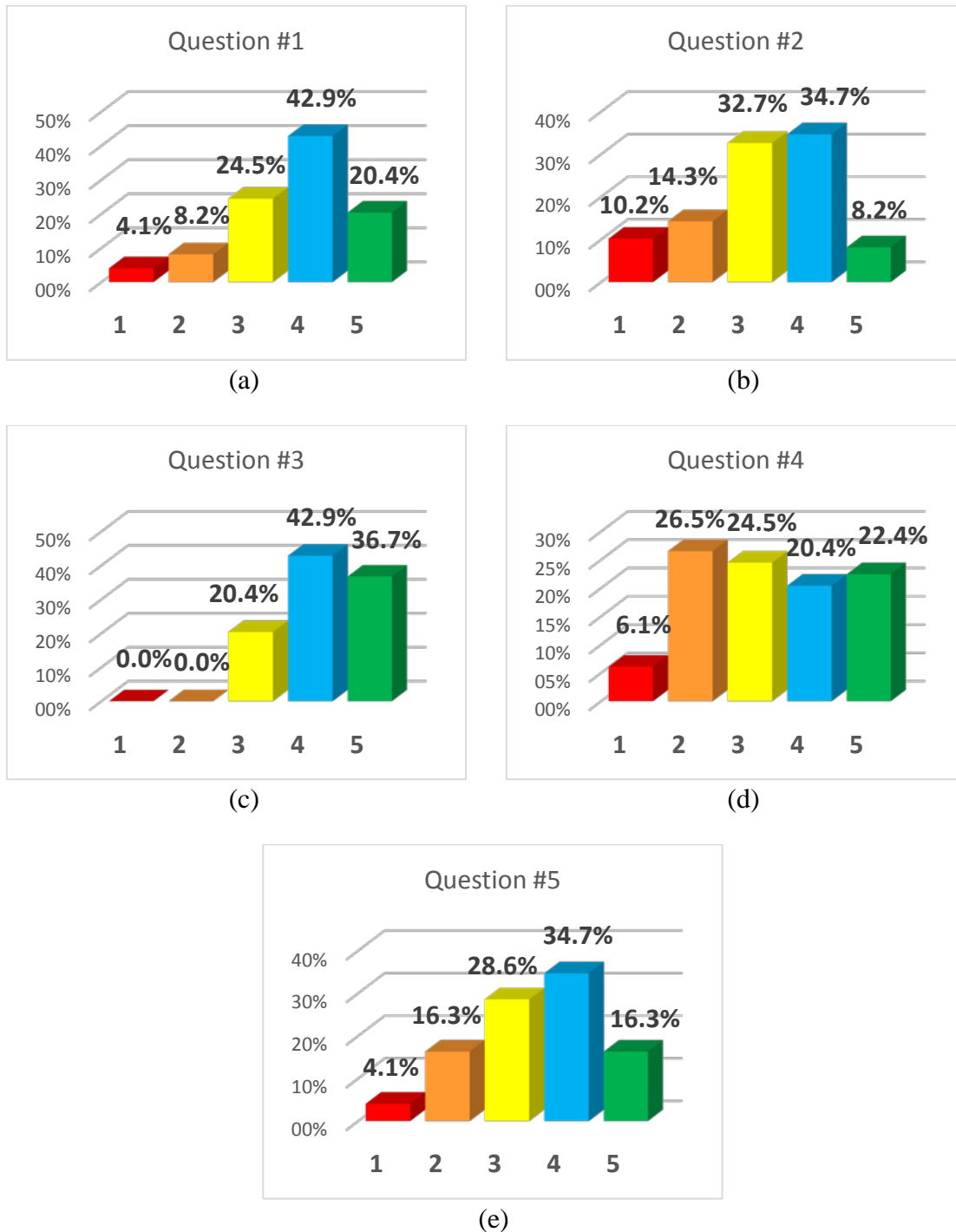


Figure 2. Results of the satisfaction survey for (a) first question, (b) second question, (c) third question, (d) fourth question and (e) fifth question.

Results of second question are also very good, showing that most of the students consider that the peer evaluation has been useful to them to better understand the subject.

Question number 3 proves that students are not reluctant to be graded by their classmates. Actually, none of them disagree on that point. On the contrary, it is usual that some students argue about the grades awarded by the professor. Therefore, we can conclude that this is also a good consequence of the methodology.

Fourth question is the one which has scored worst. Up to a 42.8% of the students have found problems in grading the works of their peers. This should be taken into account and some strategies to make this task easier in the future will have to be explored.

Finally, question number 5 show that most of the students approve the methodology, so we can conclude that it has been a good experience overall.

4. Conclusions

A peer-to-peer evaluation was carried out with engineering students. Four different sets of problems were proposed. The novelty in the methodology is that, to reduce the expected overrate in the classmates' grades, the professor randomly selected some sets of problems to be reviewed by himself and a penalty was applied if some major flaws in the students' evaluation were detected.

It has been proved that this simple strategy completely cancelled the bias. That is, the peer-to-peer and the professor's evaluation tended to be the same.

A final survey confirmed that this kind of evaluation has helped the students to keep the subject up to date and to better understand it. In addition, all of the students agreed with the grades awarded. The only point to be improved is that many students found difficult to evaluate their classmates' works but, overall, they were quite satisfied with the experience.

5. References

- Allal, L. (1991). *Vers une pratique de l'évaluation formative*. Brussels: De Boek.
- Arias Macías C.M., Arriazu Navarro R., Casanova Arias J.L., Fernández Arias J., Cárdenas Rebollo J.M. and Rey-Stolle M.F. (2014). Use of Blackboard Collaborate platform as a higher education teaching aid. *International Journal on Advances in Education Research*, 1(2), 109-124.
- Biggs, J. (2006). *Calidad del aprendizaje universitario*. [Quality of university learning.] Madrid: Narcea.
- Birembaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridway, J., Wiesemes, R. and Nickmans, G. (2006). A Learning Integrated Assessment System. *Educational Research Review*, 1(1), 61-67. <http://dx.doi.org/10.1016/j.edurev.2006.01.001>
- Bologna (1999). Joint declaration of the Ministers responsible for higher education convened in Bologna on the 19th of June. Available at http://www.magna-charta.org/resources/files/BOLOGNA_DECLARATION.pdf.
- Broadfoot, P. and Black, P. (2004). Redefining assessment? The first ten years of "Assessment in Education". *Assessment in Education*, 11(1), 7-27. <http://dx.doi.org/10.1080/0969594042000208976>
- Coll, C. and Onrubia, J. (1999). Evaluación de los aprendizajes y atención a la diversidad. In C. Coll (Coord.), *Psicología de la instrucción. La enseñanza y el aprendizaje en la educación secundaria*, (pp. 141-168). Barcelona: Horsori / ICE de la UB.
- Coll C., Rochera M.J., Mayordomo R.M., Naranjo M. (2007). Continuous assessment and support for learning: an experience in educational innovation with ICT support in higher education. *Electronic Journal of Research in Educational Psychology*, 5(3), 783-804.
- De Grez, L., Valcke, M. and Roozen, I. (2012). How Effective Are Self- and Peer Assessment of Oral Presentation Skills Compared with Teachers' Assessments? *Active*



Learning in Higher Education, 13(2), 129-142.

<http://dx.doi.org/10.1177/1469787412441284>

Delgado, A. M., Borge, R., García, J. Oliver, R. and Salomón, L. (2005). Competencias y diseño de la evaluación continua y final en el Espacio Europeo de Educación Superior. Programa de Estudios y Análisis (EA2005-0054). Madrid: Ministerio de educación y Ciencia. Dirección General de Universidades.

Díaz Barriga, F. (2006). La evaluación auténtica centrada en el desempeño: una alternativa para evaluar el aprendizaje y la enseñanza. In F. Díaz Barriga (Coord.). Enseñanza situada: vínculo entre la escuela y la vida (pp. 125-163). México: McGraw-Hill.

EHEA (2010). European Higher Education Area. Available from: <http://www.ehea.info/>.

Falchikov, N. and Goldfinch, J. Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. *Review of Educational Research*, 70(3), 287-322. <http://dx.doi.org/10.3102/00346543070003287>

Harris, J. (2011). Peer assessment in large undergraduate classes: an evaluation of a procedure for marking laboratory reports and a review of related practices. *Advances in Physiology Education*, 35(2), 178-187. <http://dx.doi.org/10.1152/advan.00115.2010>

Koç, C. (2011). The Views of Prospective Class Teachers about Peer Assessment in Teaching Practice. *Educational Sciences: Theory & Practice*, 11(4), 1979-1989.

Kommalage, M. and Gunawardena, S. (2011). *Advances in Physiology Education*, 35(1), 48-52. <http://dx.doi.org/10.1152/advan.00091.2010>

Lansiquot, R. and Rosalia C. (2015). Online Peer Review: Encouraging Student Response and Development. *Journal of Interactive Learning Research*, 26(1), 105-123.

Marks, L. and Jackson, M. (2013). Student Experience of Peer Assessment on an MSc Programme. *Bioscience Education*, 21(1), 20-28. <http://dx.doi.org/10.11120/beej.2013.00015>

Mateo Andrés, J. and Martínez Olmo, Francesc (2005). La evaluación alternativa de los aprendizajes. Cuadernos de Docencia Universitaria, nº 3, ICE-Universidad de Barcelona.

McDonald, R. (2006). The use of evaluation to improve practice in learning and teaching. *Innovations in Education and Teaching International*, 43(1), 3-13.
<http://dx.doi.org/10.1080/14703290500472087>

Nunziati, G. (1990). Pour construire un dispositif d'évaluation d'apprentissage. *Cahiers Pédagogiques*, 280, 47-64.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, and M. Zeidner (Eds.). *Handbook of self-regulation* (pp. 451-502). San Diego, CA: Academic Press. <http://dx.doi.org/10.1016/B978-012109890-2/50043-3>

Schunk, D.M. and Zimmerman, B. J. (eds.) (1998). *Self-regulated learning: From teaching to self-reflective practice*. New York: The Guilford Press.

Schlichting, K. and Fox, K. (2015). An Authentic Assessment at the Graduate Level: A Reflective Capstone Experience. *Teaching Education*, 26(3), 310-324.
<http://dx.doi.org/10.1080/10476210.2014.996748>

Torrano, F. and González, M. C. (2004). Self-Regulated Learning: Current and Future Directions. *Electronic Journal of Research in Educational Psychology*, 2(1), 1-34.

William, D. (2000). Integrating summative and formative functions of assessment. Keynote address. First Annual Conference of the European Association for Educational Assessment. Prague, Czech Republic.

Xiao, Y. and Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education*, 11(3-4), 186-193. <http://dx.doi.org/10.1016/j.iheduc.2008.06.005>